

Análise epidemiológica de HIV/AIDS no estado da Paraíba

¹Bacharelado em Ciência da Computação

Estudo de Visualização de Dados
Centro de Informática - UFPB

09 de setembro de 2025

Este trabalho tem como objetivo principal apresentar, de forma sucinta e informativa, uma análise exploratória dos casos de HIV/AIDS no estado da Paraíba durante o período de 2019 a 2024, utilizando dados provenientes do IBGE e do sistema DataSUS. Propõe-se uma análise que forneça explicabilidade aos dados, dessa forma, buscamos fornecer insights relevantes que possam auxiliar o poder público na formulação e direcionamento de ações preventivas, especialmente no contexto pós-introdução do medicamento lenacapavir, contribuindo para o bom aproveitamento de recursos públicos.

1 Escopo

Este projeto insere-se no campo científico da **Epidemiologia**, especificamente na área de epidemiologia descritiva de doenças infecciosas, com interface direta com as disciplinas de **Saúde Pública** e **Ciência da Computação**. O domínio de expertise abrange a aplicação de métodos epidemiológicos para análise de dados secundários em saúde, visando uma compreensão abrangente do impacto do HIV/AIDS na sociedade e contribuindo para o desenvolvimento de estratégias mais eficazes de prevenção, tratamento e controle da doença.

A análise epidemiológica dos casos de HIV/AIDS no estado da Paraíba constitui-se como ferramenta fundamental para identificar padrões de distribuição geográfica e demográfica, caracterizar o perfil das populações mais vulneráveis e fornecer informações essenciais para a formulação de políticas públicas direcionadas. Esta abordagem metodológica permite subsidiar ações mais efetivas e otimizar a alocação de recursos públicos limitados.

O escopo metodológico compreende a análise exploratória de dados (EDA), visualização científica de informações epidemiológicas e aplicação de princípios de epidemiologia descritiva segundo as dimensões clássicas de pessoa, lugar e tempo.

A relevância deste trabalho é reforçada pelo contexto

atual das políticas de prevenção ao HIV. Em julho de 2025, a Organização Mundial da Saúde (OMS) publicou diretrizes oficiais recomendando o uso do lenacapavir injetável como uma nova estratégia de prevenção do HIV, destacando seu potencial para ampliar o acesso e a efetividade das ações preventivas, especialmente em populações mais vulneráveis (World Health Organization, 2025). Essa recomendação internacional evidencia a necessidade de análises epidemiológicas regionais, como a proposta neste estudo, para subsidiar a implementação e o monitoramento de novas intervenções no contexto local.

2 Desafio

Do ponto de vista tecnológico, o desafio reside na integração e harmonização de múltiplas bases de dados heterogêneas (DATASUS, IBGE), que apresentem diferentes estruturas, periodicidades e níveis de granularidade. A transformação desses dados brutos em informações epidemiológicas interpretáveis e acionáveis requer o desenvolvimento de pipelines analíticos robustos e técnicas avançadas de visualização que comuniquem efetivamente padrões complexos para diferentes audiências.

O desafio operacional central é a tradução dos achados epidemiológicos em recomendações práticas para gestores de saúde pública, considerando as limitações orçamentárias e logísticas do Sistema Único de Saúde (SUS) no contexto estadual.

Adicionalmente, existe o desafio de comunicação científica: como apresentar análises epidemiológicas complexas de forma acessível e convincente para gestores públicos, profissionais de saúde e sociedade civil, garantindo que os insights gerados sejam efetivamente incorporados ao processo decisório em saúde pública.

3 Projeto de visualização

O projeto de visualização segue um modelo estruturado e envolve as seguintes etapas:

1. Entrada: foram coletadas múltiplas bases de dados epidemiológicos e demográficos, constituindo o objeto de estudo e de geração das representações visuais (RVs). Os dados epidemiológicos sobre HIV/AIDS foram extraídos do sistema DATASUS (Ministério da Saúde, 2024), abrangendo quatro bases distintas: casos por município, distribuição por sexo, faixa etária e valores de internação. Os dados demográficos da Paraíba foram obtidos do IBGE (Instituto Brasileiro de Geografia e Estatística, 2024b), incluindo informações populacionais do censo 2022, além da malha territorial através de shapefile (Instituto Brasileiro de Geografia e Estatística, 2024a). Todas as bases foram importadas através da biblioteca pandas.

2. Pré-Processamento: foram realizadas transformações essenciais para limpeza e harmonização dos dados, incluindo padronização de nomes de municípios, tratamento de valores ausentes e cálculo da taxa de incidência por 100.000 habitantes.

3. Mapeamento: foi estabelecido um projeto gráfico baseado em questionamentos epidemiológicos fundamentais, definindo paleta de cores e estruturas visuais adequadas para representar distribuições geográficas, temporais e demográficas dos casos de HIV/AIDS.

4. Visualização: etapa final que apresenta as RVs através de mapas da distribuição de casos por município, análises temporais da evolução epidemiológica e caracterização demográfica por sexo e faixa etária, subsidiando a tomada de decisão em saúde pública.

4 Mapeamento

O tratamento dos dados (codificação resumida) foi realizado sobre as bases de dados epidemiológicos do DATASUS e demográficos do IBGE, que por sua vez foram processadas através de técnicas de ETL (Extrair, Transformar e Carregar) utilizando Python.

Processamento das Bases DATASUS

Os códigos abaixo sintetizam o tratamento realizado para limpeza e união das quatro bases epidemiológicas em dataframes estruturados:

```
1 import pandas as pd
2
3 # Importacao das bases DATASUS e IBGE
4 df_mun = pd.read_csv('datasus_hiv_mun_2019
5 -2024.csv', sep=';')
6 df_sex = pd.read_csv('
7 datasus_hiv_sex_data_2019-2024.csv', sep=
8 ';')
9 df_sex_age = pd.read_csv('
10 datasus_hiv_sex_idade_2019-2024.csv', sep
11 =';')
```

```
7 df_mun_val = pd.read_csv('
8 datasus_hiv_mun_valor_2019-2024.csv', sep
9 =';')
10 df_pop = pd.read_excel("ibge_pb_mun.xlsx")
11
12 # Limpeza e padronizacao
13 df_mun['Municipio'] = df_mun['Municipio'].
14 astype(str).str.replace(r'\d+s+', '',
15 regex=True)
16 df_mun_val['Municipio'] = df_mun_val['
17 Municipio'].astype(str).str.replace(r'\d
18 +s+', '', regex=True)
```

União com Dados Demográficos

O IBGE forneceu informações populacionais que, após união com os dados epidemiológicos (código abaixo), produziram o dataframe para cálculo de taxas de incidência:

```
1 # Padronizacao para merge
2 df_pop['Municipio'] = df_pop['Municipio'].
3 apply(
4 lambda x: remove_accents(str(x)).upper())
5
6 # Uniao das bases
7 df_final = df_mun_2022.merge(df_pop, on='
8 Municipio')
9 df_final['Taxa_100mil'] = (df_final['Casos']
10 /
11 df_final['
12 Populacao']) *
13 100000
14
15 # Filtro para municipios com populacao >=
16 10.000 hab
17 df_final = df_final[df_final['Populacao'] >=
18 10000]
```

A Figura 1 apresenta o resultado do processo de união das bases epidemiológicas com os dados populacionais, exemplificando a estrutura final dos dados após o tratamento, incluindo os municípios paraibanos com suas respectivas distribuições mensais de casos, população e taxa calculada por 100.000 habitantes.

	Município	2022/Jan	2022/Fev	2022/Mar	2022/Abr	2022/Mai	2022/Jun	2022/Jul	2022/Ago	2022/Set	2022/Oct	2022/Nov	2022/Dez	Total	População	Casos por 100mil habitantes
2	BABELOX	24	9	9	8	11	19	14	12	11	8	11	7	143	6274210	172,85
3	CABEDLO	8	9	7	7	21	10	6	16	14	2	1	2	109	4601630	155,50
4	JOÃO PESSOA	72	43	188	46	121	77	101	76	94	100	24	14	554	6336130	114,46
5	SANTA RITA	6	5	13	12	16	15	15	7	16	10	8	15	154	14991020	69,35
6	MAMANGUAPE	1	3	1	3	4	2	5	3	5	2	3	6	37	4409108	62,55

Figura 1: Dataframe pós merge

Destaca-se que na construção da análise de taxa por 100.000 habitantes foram desconsiderados os casos de municípios com menos de 10.000 habitantes, visto que a baixa população poderia gerar inconsistências na análise, resultando em taxas artificialmente elevadas que não representariam adequadamente o cenário epidemiológico real.

No mais, os dados geoespaciais foram adquiridos através do shapefile do IBGE, permitindo a criação de mapas coropléticos para visualização da distribuição geográfica dos casos de HIV/AIDS no estado da Paraíba.

5 Representações Visuais

As RVs estão dispostas a seguir, seguindo todo o processo mencionado no projeto de visualização .

Evolução Temporal dos Casos

A Figura 2 representa a progressão anual dos casos de HIV/AIDS na Paraíba. Do ponto de vista visual, pode-se notar que as barras possuem coloração uniforme em tom azul, facilitando a comparação direta entre os valores anuais. Observa-se também a presença de anotações percentuais acima de cada barra, mecanismo que quantifica as variações ano a ano, e uma linha horizontal tracejada indicando a média do período, elemento que serve como referência visual para identificação de desvios da tendência central.

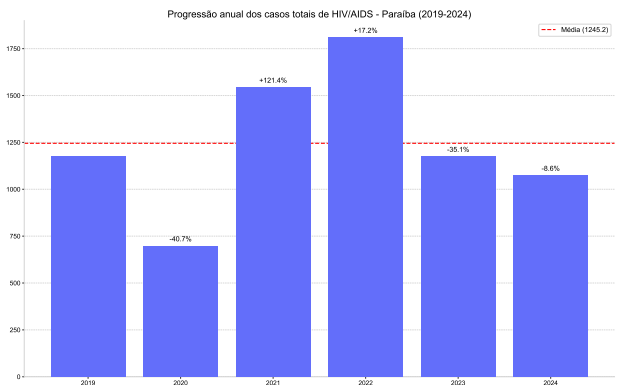


Figura 2: Evolução temporal dos casos de HIV/AIDS na Paraíba (2019-2024)

Distribuição por Sexo

A Figura 3, por sua vez, indica a distribuição temporal estratificada por sexo. Em relação à visualização, adotou-se o gráfico de linhas com diferenciação cromática: azul para o sexo masculino e rosa para o feminino, seguindo convenções visuais estabelecidas.

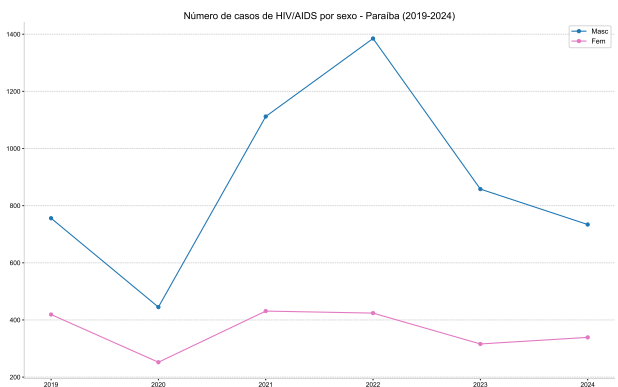


Figura 3: Distribuição de casos de HIV/AIDS por sexo na Paraíba (2019-2024)

Casos por Faixa Etária e Sexo

A Figura 4 aponta a distribuição dos casos segundo faixa etária estratificada por sexo. Nesta visualiza-

ção, foi adotada uma composição de barras agrupadas com diferenciação cromática entre os sexos: azul para masculino e laranja para feminino. Pode-se notar também a presença de dois subgráficos lado a lado, sendo o primeiro com barras em tons de cinza para visualização geral e o segundo com as barras agrupadas coloridas, facilitando a comparação entre os sexos dentro de cada faixa etária.

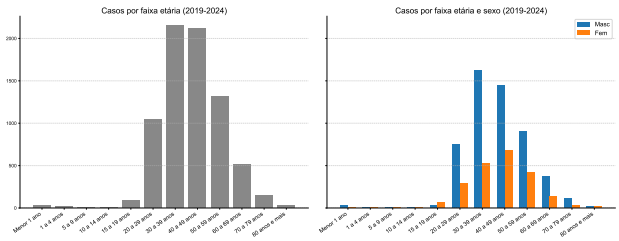


Figura 4: Casos de HIV/AIDS por faixa etária e sexo na Paraíba (2019-2024)

Distribuição Geográfica Municipal

A Figura 5, por sua vez, representa a distribuição espacial através de mapa coroplético. Do ponto de vista visual, adotou-se uma escala de cores em gradiente viridis, onde tonalidades mais escuras indicam maiores concentrações de casos por 100.000 habitantes. Observa-se também a presença de uma barra de legenda lateral, facilitando a interpretação quantitativa das cores, e contornos municipais bem definidos em preto, conferindo clareza na delimitação territorial.

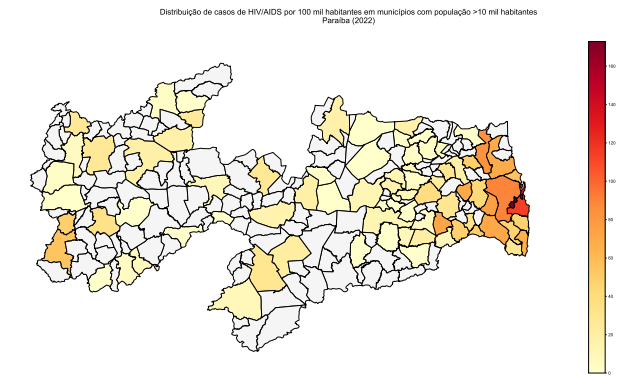


Figura 5: Distribuição geográfica de casos (2022)

Ranking Municipal por Taxas

Por fim, a Figura 6 apresenta o ranking dos municípios com maiores taxas de incidência. Em relação à visualização, utilizou-se um gráfico de barras horizontais em gradiente de cores, variando de tons mais escuros (azul-violeta) para mais claros (verde-amarelo) conforme a magnitude das taxas. Pode-se verificar também a presença de rótulos municipais rotacionados, mecanismo que facilita a leitura e identificação dos municípios prioritários.

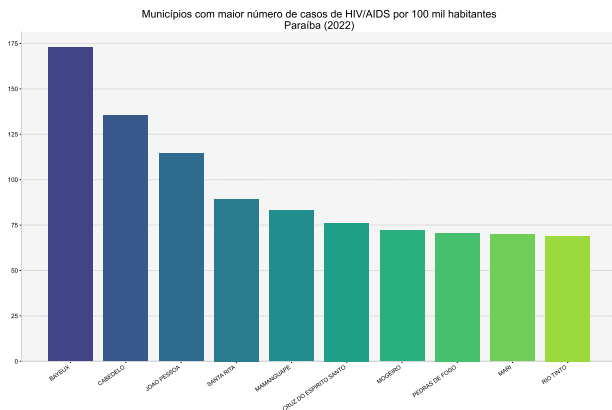


Figura 6: Municípios da Paraíba com maiores taxas de HIV/AIDS por 100.000 habitantes (2022)

6 Análise de dados

A partir das representações visuais desenvolvidas, observa-se que o perfil epidemiológico do HIV/AIDS na Paraíba apresenta características bem definidas. O grupo mais vulnerável concentra-se na população masculina entre 30-49 anos, representando aproximadamente 40% dos 7.471 casos registrados no período 2019-2024. A distribuição geográfica dos 1.809 casos de 2022 revela concentração na região metropolitana de João Pessoa, com Bayeux apresentando a maior taxa (172,8 casos/100.000 hab.), seguida pelos municípios litorâneos de Cabedelo e

Santa Rita.

A análise temporal evidencia impacto significativo da pandemia de COVID-19, com redução de 40,7% dos casos em 2020, seguida de recuperação expressiva em 2021-2022. A estratificação por sexo mantém padrão consistente de predominância masculina (razão 2,5:1), enquanto a distribuição etária confirma que adultos jovens constituem o grupo de maior risco, demandando estratégias preventivas direcionadas.

7 Takeaways

1. Perfil de Risco Definido: Homens de 30-49 anos representam o grupo mais vulnerável ao HIV/AIDS na Paraíba, concentrando a maioria dos casos e demandando campanhas preventivas específicas para este segmento populacional.

2. Concentração Geográfica Estratégica: A região metropolitana de João Pessoa, especialmente os municípios de Bayeux, Cabedelo e Santa Rita, apresenta as maiores taxas de incidência, indicando a necessidade de intensificação dos serviços de testagem e prevenção nestas localidades.

3. Impacto de Fatores Externos: A pandemia de COVID-19 alterou significativamente os padrões de diagnóstico, evidenciando a importância de sistemas de vigilância epidemiológica robustos e resilientes para garantir continuidade do monitoramento e controle da doença.

8 Referências

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Malhas Territoriais**. [S. l.: s. n.], 2024. <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html>. Acessado em: 09 set. 2025.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Paraíba | Cidades e Estados**. [S. l.: s. n.], 2024. <https://www.ibge.gov.br/cidades-e-estados/pb.html>. Acessado em: 09 set. 2025.

MINISTÉRIO DA SAÚDE. **DATASUS - Informações de Saúde (TABNET)**. [S. l.: s. n.], 2024. <https://datasus.saude.gov.br/informacoes-de-saude-tabnet>. Acessado em: 09 set. 2025.

WORLD HEALTH ORGANIZATION. **Guidelines on lenacapavir for HIV prevention and testing strategies for long-acting injectable preexposure prophylaxis (PrEP)**. Geneva, 2025. Licence: CC BY-NC-SA 3.0 IGO.